

Compressive Channel Estimation and Multi-user Detection in C-RAN

Qi He^{†§}, Tony Q.S. Quek[§], Zhi Chen[†] and Shaoqian Li[†]

[†] National Key Laboratory on Communications, University of Electronic Science and Technology of China

[§] Information Systems Technology and Design Pillar, Singapore University of Technology and Design

heqi.tech@hotmail.com; tonyquek@sutd.edu.sg; chenzhi@uestc.edu.cn; lsq@uestc.edu.cn

Abstract—This paper considers the channel estimation (CE) and multi-user detection (MUD) problems in cloud radio access network (C-RAN). Assuming that active users are sparse in the network, we solve CE and MUD problems with compressed sensing (CS) technology to greatly reduce the long identification pilot overhead. A mixed $\ell_{2,1}$ -regularization functional for extended sparse group-sparsity recovery is proposed to exploit the inherently sparse property existing both in user activities and remote radio heads (RRHs) that active users are attached to. Empirical and theoretical guidelines are provided to help choosing tuning parameters which have critical effect on the performance of the penalty functional. To speed up the processing procedure, based on alternating direction method of multipliers and variable splitting strategy, an efficient algorithm is formulated which is guaranteed to be convergent. Numerical results are provided to illustrate the effectiveness of the proposed functional and efficient algorithm.

I. INTRODUCTION

The cloud radio access network (C-RAN) has been proposed as a breakthrough network architecture to improve spectral efficiency and energy efficiency [1], [2]. In C-RAN, strong functional but high-power-consumption base stations (BSs) are decoupled into two parts: distributed remote radio heads (RRHs) and baseband units (BBUs) [3]. The BBUs are consolidated together to be a BBU pool for handling all the baseband signal processing in the network. Plenty of RRHs are distributed on a large area, and connected to BBU pool through fronthaul links.

The estimation of channel state information (CSI) is fundamental and vital in wireless communication. The identification of active users is also critical for resource allocation in the network. Channel estimation (CE) and multi-user detection (MUD) problems have been investigated extensively [4]. However, in classical CE and MUD methods, in order to keep orthogonal characteristic between identification pilots belong to different users, the length of identification pilots needs to be scale with the number of users times the number of antennas per user, which is a significant overhead when the network is large. With the observation that active users are sparse, many recent works [5]–[8] on MUD and CE have taken the sparsity of user activities into consideration by using compressed sensing (CS) technology, which greatly reduces the identification pilot overhead.

The C-RAN architecture brings new challenges to CE and MUD. Since the baseband processing is clustered together in C-RAN, CE and MUD is no longer restricted to local BS processing, but jointly processed in BBU pool to meet the demand of the whole network. Each user is linked to each RRH to enable full cooperation among RRHs. This paper

aims to detect the active users and to estimate full CSI, i.e. the channel states between each active user and each RRH in C-RAN.

Related works on CE and MUD in C-RAN can be found in [9] and [10]. Bayesian sparsity inference tool and hybrid generalized approximate message passing method are leveraged in [9] and [10] respectively to exploit sparsity of user activities. [9] assumes multiple antennas in each RRH and user, while [10] adopts single-antenna RRHs and users but takes limited capacity of fronthaul links into consideration. Both [9] and [10] assume the Rayleigh fading channel and the exact knowledge of large-scale path loss. However, this simplified model tends to describe static network, while mobility and environment changes introduce more uncertainty [11], [12]. Besides, to predict the precise large-scale path loss between each RRH and each user in C-RAN is very expensive when the network is large [13], [14].

In this paper, we formulate a penalty functional to solve CE and MUD problems in C-RAN. Our formulation is general which does not depend on the prior information of channel parameters or the assumption of channel model. We only assume the sparsity of active users, which is practical in real communication systems. On the other hand, we observe that there are limited RRHs surrounding each active user in C-RAN, and the effects of most distant RRHs on this user are negligible. Hence there exist two types of sparsity: one is the sparse active users; the other is sparse RRHs related to one particular active user. Motivated by this observation, we propose a mixed $\ell_{2,0}$ -regularization functional to exploit these two types of sparsity simultaneously, and then relax it as a convex and smoothed mixed $\ell_{2,1}$ -regularization functional. A re-weighting strategy is adopted to enhance estimation accuracy. As tuning parameters in the functional have critical influence on the performance, we also provide empirical and theoretical guidelines for choosing the appropriate values.

Considering that the processing time of MUD and CE should be less than the channel coherence time, we propose an efficient algorithm based on alternating direction method of multipliers (ADMM) and variable splitting strategy to reduce the computational complexity of solving the proposed functional. Simulation results show that, even without any prior information on the channel, our algorithm provides almost the same performance and has almost the same computational complexity as state-of-the-art algorithm.

Notations: Uppercase (lowercase) boldface letters denote matrices (column vectors). *iff* denotes “if and only if”. The operators $(\cdot)^T$, $(\cdot)^H$, $\|\cdot\|_F$ stand for transpose, conjugate

transpose and Frobenius norm, respectively. \mathbf{I}_N denotes an $N \times N$ identity matrix. \otimes and \circ denote the Kronecker product and Hadamard product (element product) respectively. $\Re(a)$ stands for real part of complex number a . $\text{vec}(\mathbf{A})$ denotes the vectorization of \mathbf{A} formed by stacking its columns into a single column vector. $\mathbf{A}_{[i,j]}$ stands for the $\{i, j\}$ th element in \mathbf{A} , and $\mathbf{A}_{i,j}$ for the $\{i, j\}$ th submatrix in \mathbf{A} . $\mathbf{A}^{\circ(a)}$ denotes an element-wise exponentiation operator on \mathbf{A} , i.e., $(\mathbf{A}^{\circ(a)})_{[i,j]} = (\mathbf{A}_{[i,j]})^a$.

II. SYSTEM MODEL

We consider a C-RAN architecture with G RRHs and K users. There are M antennas in each RRH, and N antennas in each user device. The length of identification pilots is L , and the k th user is assigned with a pilot matrix $\mathbf{P}_k \in \mathbb{C}^{N \times L}$. We assume that only a small part of users are active. The set of active users is indicated by $\mathcal{A} \subseteq \{1, \dots, K\}$. An indicator function is defined to mark whether the k th user is active,

$$\mathbb{1}_{\mathcal{A}}(k) = \begin{cases} 1, & \text{if } k \in \mathcal{A} \\ 0, & \text{if } k \notin \mathcal{A} \end{cases} \quad (1)$$

In the following descriptions, we use $\mathbb{1}_k$ to denote $\mathbb{1}_{\mathcal{A}}(k)$ if there is no confusion.

The quasi-static channel between k th user and g th RRH is denoted by complex matrix $\mathbf{H}_{g,k} \in \mathbb{C}^{M \times N}$. As only active users transmit their identification pilots, the received pilot data in the g th RRH can be described as

$$\mathbf{R}_g = \sum_{k \in \mathcal{A}} \mathbf{H}_{g,k} \mathbf{P}_k + \bar{\mathbf{N}}_g = \sum_{k=1}^K \mathbf{H}_{g,k} \mathbb{1}_k \mathbf{P}_k + \bar{\mathbf{N}}_g, \quad g = 1, \dots, G \quad (2)$$

where $\bar{\mathbf{N}}_g \in \mathbb{C}^{M \times L}$ denotes the antenna additive noise in the g th RRH.

By stacking received pilot data in all the G RRHs, the data received in BBU can be described as

$$\mathbf{R} = \mathbf{H} \mathbf{A} \mathbf{P} + \bar{\mathbf{N}} \quad (3)$$

where $\mathbf{R} = [\mathbf{R}_1^T, \dots, \mathbf{R}_G^T]^T$, $\mathbf{A} = \text{diag}[\mathbb{1}_1, \dots, \mathbb{1}_K] \otimes \mathbf{I}_N$, $\mathbf{P} = [\mathbf{P}_1^T, \dots, \mathbf{P}_K^T]^T$, $\bar{\mathbf{N}} = [\bar{\mathbf{N}}_1^T, \dots, \bar{\mathbf{N}}_G^T]^T$, and

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{1,1} & \cdots & \mathbf{H}_{1,K} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{G,1} & \cdots & \mathbf{H}_{G,K} \end{bmatrix} \quad (4)$$

Besides the received pilot data \mathbf{R} , BBU also has the knowledge of identification pilot matrix \mathbf{P} . Hence, our MUD problem is to estimate \mathbf{A} in (3). With the constraint of sparse user activities, our CE problem is to estimate matrices $[\mathbf{H}_{1,k}^T; \dots; \mathbf{H}_{G,k}^T]^T$, $\forall k \in \mathcal{A}$.

III. PROBLEM FORMULATIONS

By taking the transposition of equation (3), we get a compressed sensing problem as below.

$$\mathbf{R}^H = \mathbf{P}^H \mathbf{A} \mathbf{H}^H + \bar{\mathbf{N}}^H \quad (5)$$

where \mathbf{P}^H is a fat matrix with the assumption that the pilot length L is less than the number of users K . Our MUD and CE problems are converted to estimating $\mathbf{A} \mathbf{H}^H$ as a whole.

$$\mathbf{A} \mathbf{H}^H = \begin{bmatrix} \mathbb{1}_1 \mathbf{H}_{1,1}^H & \cdots & \mathbb{1}_1 \mathbf{H}_{G,1}^H \\ \vdots & \ddots & \vdots \\ \mathbb{1}_K \mathbf{H}_{1,K}^H & \cdots & \mathbb{1}_K \mathbf{H}_{G,K}^H \end{bmatrix} \quad (6)$$

We define the submatrix $[\mathbb{1}_k \mathbf{H}_{1,k}^H, \dots, \mathbb{1}_k \mathbf{H}_{G,k}^H]$ as the k th “row chunk” matrix in $\mathbf{A} \mathbf{H}^H$, and the submatrix $\mathbb{1}_k \mathbf{H}_{g,k}^H$ as the $\{k, g\}$ th “element chunk” matrix. The sparsity of the k th row chunk is determined by $\mathbb{1}_k$, which denotes whether the k th user is active. Therefore the matrix $\mathbf{A} \mathbf{H}^H$ has row-chunk sparsity structure at first.

In practical C-RAN architecture, huge numbers of RRHs are distributed on vast areas, and most of RRHs have negligible effect on a particular active user. It gives rise to that, if the k th user is active, most of the element chunks in the k th row chunk approximate to zero, and the k th row chunk has sparsity structure. Therefore, we say $\mathbf{A} \mathbf{H}^H$ has element-chunk sparsity structure. For one particular active user, the element-chunk sparsity situation is affected by many factors, like the position of this user, the distribution geometry of surrounding RRHs, power allocations and beamformings in RRHs, etc.

To simplify the expression, we rewrite the system model (5) as below,

$$\mathbf{B} = \mathbf{A} \mathbf{X} + \mathbf{N} \quad (7)$$

where $\mathbf{B} = \mathbf{R}^H \in \mathbb{C}^{L \times GM}$, $\mathbf{A} = \mathbf{P}^H \in \mathbb{C}^{L \times KN}$, $\mathbf{X} = \mathbf{A} \mathbf{H}^H \in \mathbb{C}^{KN \times GM}$, and $\mathbf{N} = \bar{\mathbf{N}}^H$.

The matrix \mathbf{X} to be estimated has both row-chunk and element-chunk sparsity structures. In the following, we use $\mathbf{X}_i \in \mathbb{C}^{N \times GM}$ to denote the i th row chunk $[\mathbb{1}_k \mathbf{H}_{1,k}^H, \dots, \mathbb{1}_k \mathbf{H}_{G,k}^H]$, and $\mathbf{X}_{i,j} \in \mathbb{C}^{N \times M}$ to denote the $\{i, j\}$ th element chunk $\mathbb{1}_i \mathbf{H}_{j,j}^H$. We define $\mathbf{A}_i = \mathbf{P}_i^T \in \mathbb{C}^{L \times N}$ in (5), then $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_K]$; and define $\mathbf{B}_i = \mathbf{R}_i^T \in \mathbb{C}^{L \times M}$ in (5), then $\mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_G]$.

To exploit row-chunk and element-chunk sparsity structures simultaneously, we propose a mixed $\ell_{2,0}$ -regularization functional shown as below,

$$\min_{\mathbf{X}} \alpha_1 \sum_{i=1}^K \mathbb{1}(\mathbf{X}_i) + \alpha_2 \sum_{i=1}^K \sum_{j=1}^G \mathbb{1}(\mathbf{X}_{i,j}) + \frac{1}{2} \|\mathbf{A} \mathbf{X} - \mathbf{B}\|_F^2 \quad (8)$$

where $\mathbb{1}(\cdot)$ is an indicator function defined as following,

$$\mathbb{1}(\mathbf{D}) = \begin{cases} 1, & \text{if } \mathbf{D} \text{ is a non-zero matrix} \\ 0, & \text{if } \mathbf{D} \text{ is a zero matrix} \end{cases} \quad (9)$$

In the functional (8), the first term exploits the row chunk sparsity, and the second term takes into account the element chunk sparsity. The absolute values of tuning parameters α_1 and α_2 control the tradeoff between the sparsity of the solution and the quality of fit, while the relative values between α_1 and α_2 control the balance of row-chunk and element-chunk sparsity.

As (8) is non-smooth, we will relax it to a smooth and convex mixed $\ell_{2,1}$ -regularization functional. However, there is one drawback in $\ell_{2,1}$ -regularization functionals that coefficients with larger magnitudes are penalized more heavily than coefficients with smaller magnitudes, which limits the performance and also exhibits a key difference between $\ell_{2,1}$ and $\ell_{2,0}$ regularizations.

An iteratively re-weighted ℓ_1 -norm minimization algorithm was proposed in [15] to provide approximative “democratic penalization” on coefficients and further enhance the performance of ℓ_1 -norm functional. Similarly to that, we

introduce the weight matrix \mathbf{W} , and propose a weighted $\ell_{2,1}$ -regularization functional as below,

$$\min_{\mathbf{X}} \alpha_1 \sum_{i=1}^K \|\mathbf{W}_i \circ \mathbf{X}_i\|_F + \alpha_2 \sum_{i=1}^K \sum_{j=1}^G \|\mathbf{W}_{i,j} \circ \mathbf{X}_{i,j}\|_F + \frac{1}{2} \|\mathbf{AX} - \mathbf{B}\|_F^2 \quad (10)$$

where \mathbf{W}_i and $\mathbf{W}_{i,j}$ denote the i th row-chunk submatrix and the $\{i, j\}$ th element-chunk submatrix in \mathbf{W} respectively.

The functional (10) is in the context of multiple measurement vectors (MMV) scenario. If we suppose that each RRH and user owns only one antenna, i.e., $M = N = 1$, then each element chunk $\mathbf{X}_{i,j}$ reduces to a complex number. By letting $\mathbf{b} = \text{vec}(\mathbf{B}^T)$, $\mathbf{x} = \text{vec}(\mathbf{X}^T)$, $\mathbf{n} = \text{vec}(\mathbf{N}^T)$ and $\hat{\mathbf{A}} = \mathbf{A} \otimes \mathbf{I}_G$, from equation (7) we get

$$\mathbf{b} = \hat{\mathbf{A}}\mathbf{x} + \mathbf{n} \quad (11)$$

which is a single measurement vector (SMV) model. The vector \mathbf{x} to be estimated has two types of sparsity: “group-wise sparsity” and “element-wise sparsity within group”, where the group size is G . By removing the effect of weight matrix \mathbf{W} , our proposed functional (10) is reduced to a sparse group lasso criterion [16] as following,

$$\min_{\mathbf{x}} \alpha_1 \sum_{i=1}^K \|\mathbf{x}^{(i)}\|_F + \alpha_2 \|\mathbf{x}\|_1 + \frac{1}{2} \|\hat{\mathbf{A}}\mathbf{x} - \mathbf{b}\|_F^2 \quad (12)$$

where $\mathbf{x}^{(i)}$ denotes the i th group in \mathbf{x} , and $\|\mathbf{x}\|_1$ denotes the sum of magnitudes of each element in \mathbf{x} .

Considering feasibility of multiple antennas in RRHs and users, and also higher estimation accuracy demand, we adopt the functional (10), which can be viewed as extended version of sparse group-sparsity recovery problem. If only single-antenna RRHs and user devices are deployed in the C-RAN architecture, (12) can be used to get normal precise estimation, and can be solved by several existing algorithms [16], [17], etc.

Next, we discuss the choices of α_1 and α_2 , which have critical effect on the performance of (10). Smaller values of α_1 and α_2 may be insufficient to recover the interested sparse signal, and larger values may lead to biased estimation. However, to the best of our knowledge, determining proper values for tuning parameters like α_1 and α_2 still remains an implementation-level issue.

Remark 1. The choice of α_1 and α_2 are empirical. In general, a larger α_1 should be chosen when the C-RAN is small and dense, where the row-chunk sparsity dominates and the element-chunk sparsity is non-significant. When the C-RAN is large, we empirically choose a larger α_2 to exploit the element-chunk sparsity efficiently.

Here we further provide rigorously numerical upper bounds to help choosing α_1 and α_2 . We notice that if $\alpha_1 = 0$ or $\alpha_2 = 0$, only one type sparsity is grasped, and (10) reduces to an extension of group-lasso functional [18]. Inspired by this insight, we further arrive at the following result.

Theorem 1. The solution of the functional (10) is $\hat{\mathbf{X}} = \mathbf{0}$ if $\alpha_1 \geq \alpha_1^* = \max_i \|(\mathbf{A}_i^H \mathbf{B}) \circ (\mathbf{W}_i)^{\circ(-1)}\|_F$ or $\alpha_2 \geq \alpha_2^* = \max_{i,j} \|(\mathbf{A}_i^H \mathbf{B}_j) \circ (\mathbf{W}_{i,j})^{\circ(-1)}\|_F$.

The proof of Theorem 1 can be found in Appendix A. It can be seen that α_1 should be chosen strictly less than α_1^* , and α_2 less than α_2^* , to prevent identically zero solution.

IV. ALGORITHM DESIGN

A. Standard Second-Order Cone Programming

The proposed functional (10) can be transformed into a second-order cone programming (SOCP) problem as below, and solved by standard method, such as interior point method [19].

$$\begin{aligned} \min_{\mathbf{X}} \quad & \alpha_1 \sum_{i=1}^K \mathbf{c}_i + \alpha_2 \sum_{i=1}^K \sum_{j=1}^G \mathbf{D}_{i,j} + e \\ \text{s.t.} \quad & \|\mathbf{W}_i \circ \mathbf{X}_i\|_F \leq \mathbf{c}_i \\ & \|\mathbf{W}_{i,j} \circ \mathbf{X}_{i,j}\|_F \leq \mathbf{D}_{i,j} \\ & \frac{1}{2} \|\mathbf{AX} - \mathbf{B}\|_F^2 \leq e \end{aligned} \quad (13)$$

where $\mathbf{c} \in \mathbb{R}_+^{K \times 1}$ and $\mathbf{D} \in \mathbb{R}_+^{K \times G}$.

The SOCP method is guaranteed to find the optimal solution of our convex functional (10) with fixed \mathbf{W} . We iterate the procedure (13) for several times, in each of which we use previous estimation of \mathbf{X} to determine \mathbf{W} similarly to [15]. As an empirical law, each element in \mathbf{W} is set inversely to previous estimation of element magnitude, as below

$$\mathbf{W}_{[i,j]} = \frac{1}{\text{abs}(\mathbf{X}_{[i,j]}) + \epsilon} \quad (14)$$

where $\epsilon > 0$ is a very small positive number to provide stability.

B. Alternating Direction Method of Multipliers

Solving SOCP problem in (13) has high computational complexity when the problem size is medium to large scale. In practical C-RAN system, the huge number of users and RRHs make it impractical to solve (10) by standard algorithm like SOCP. Given that, we propose an efficient algorithm to accelerate the problem-solving procedure. The alternating direction method of multipliers (ADMM) combined with variable splitting strategy can be used to separate our problem into simple sub-problems. Methods based on ADMM are easy to implement and also own a guaranteed convergency property [20].

We first split the variable \mathbf{X} in (10), and bring two auxiliary variables \mathbf{Z} and \mathbf{Q} :

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{Z}, \mathbf{Q}} \quad & \alpha_1 \sum_{i=1}^K \|\mathbf{Z}_i\|_F + \alpha_2 \sum_{i=1}^K \sum_{j=1}^G \|\mathbf{Q}_{i,j}\|_F + \frac{1}{2} \|\mathbf{AX} - \mathbf{B}\|_F^2 \\ \text{s.t.} \quad & \mathbf{Z} = \mathbf{W} \circ \mathbf{X}, \quad \mathbf{Q} = \mathbf{W} \circ \mathbf{X} \end{aligned} \quad (15)$$

As (15) is separable in variables \mathbf{Z} , \mathbf{Q} and \mathbf{X} , ADMM is applicable. We transform (15) to be an augmented Lagrangian problem as below,

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{Q}, \lambda_1, \lambda_2) \\ = \alpha_1 \sum_{i=1}^K \|\mathbf{Z}_i\|_F + \alpha_2 \sum_{i=1}^K \sum_{j=1}^G \|\mathbf{Q}_{i,j}\|_F + \frac{1}{2} \|\mathbf{AX} - \mathbf{B}\|_F^2 \\ - \Re \left\{ \text{tr} \left(\lambda_1^H (\mathbf{Z} - \mathbf{W} \circ \mathbf{X}) \right) + \text{tr} \left(\lambda_2^H (\mathbf{Q} - \mathbf{W} \circ \mathbf{X}) \right) \right\} \\ + \frac{\beta}{2} (\|\mathbf{Z} - \mathbf{W} \circ \mathbf{X}\|_F^2 + \|\mathbf{Q} - \mathbf{W} \circ \mathbf{X}\|_F^2) \end{aligned} \quad (16)$$

where λ_1 and λ_2 are Lagrange multipliers which have the same dimension as \mathbf{X} . β is the regularization parameter.

The ADMM iterates over the minimization of augmented Lagrangian problem (16) on variables \mathbf{X} , \mathbf{Z} and \mathbf{Q} sequentially, each of which forms a subproblem as below.

The \mathbf{X} -subproblem is to minimize (16) with respect to \mathbf{X} , which is a convex quadratic problem. We have the following equation by taking derivative with respect to \mathbf{X} and setting it to be zero.

$$\begin{aligned} \beta \mathbf{W} \circ (\mathbf{Z} + \mathbf{Q}) + \mathbf{A}^H \mathbf{B} - \mathbf{W} \circ (\lambda_1 + \lambda_2) \\ = 2\beta \mathbf{W}^{o(2)} \circ \mathbf{X} + \mathbf{A}^H \mathbf{A} \mathbf{X} \end{aligned} \quad (17)$$

By setting the left side in (17) equal to a newly defined matrix \mathbf{D} , we get the estimation of the l th column in \mathbf{X} ,

$$\begin{aligned} \hat{\mathbf{X}}_{[:,l]} &= \left(2\beta \text{diag}((\mathbf{W}_{[:,l]})^{o(2)}) + \mathbf{A}^H \mathbf{A} \right)^{-1} \mathbf{D}_{[:,l]} \\ &= \left(\mathbf{P}_l - \mathbf{P}_l \mathbf{A}^H (\mathbf{I}_L + \mathbf{A} \mathbf{P}_l \mathbf{A}^H)^{-1} \mathbf{A} \mathbf{P}_l \right) \mathbf{D}_{[:,l]} \end{aligned} \quad (18)$$

for $l = 1, \dots, GM$, where $\mathbf{P}_l := (0.5/\beta) \text{diag}((\mathbf{W}_{[:,l]})^{o(-2)})$.

The \mathbf{Z} -subproblem is to minimize (16) with respect to \mathbf{Z} , which is given by

$$\min_{\mathbf{Z}} \alpha_1 \sum_{i=1}^K \|\mathbf{Z}_i\|_F - \Re\{tr(\lambda_1^H \mathbf{Z})\} + \frac{\beta}{2} \|\mathbf{Z} - \mathbf{W} \circ \mathbf{X}\|_F^2 \quad (19)$$

Clearly, the minimization problem (19) can be solved separately for each row chunk \mathbf{Z}_i . With a little algebra, (19) is equivalent to

$$\sum_{i=1}^K \min_{\mathbf{Z}_i} \left[\alpha_1 \|\mathbf{Z}_i\|_F + \frac{\beta}{2} \left\| \mathbf{Z}_i - \mathbf{W}_i \circ \mathbf{X}_i - \frac{1}{\beta_1} (\lambda_1)_i \right\|_F^2 \right] \quad (20)$$

For the i th subproblem in (20), we show in Appendix B that it has a closed-form solution. Then the optimal solution of (20) is

$$\hat{\mathbf{Z}}_i = \max \left\{ \left\| \tilde{\mathbf{Z}}_i \right\|_F - \frac{\alpha_1}{\beta}, 0 \right\} \frac{\tilde{\mathbf{Z}}_i}{\left\| \tilde{\mathbf{Z}}_i \right\|_F}, \quad \forall i = 1, \dots, K, \quad (21)$$

where $\tilde{\mathbf{Z}} = \mathbf{W} \circ \mathbf{X} + (1/\beta) \lambda_1$. To simplify expression, we denote the above chunk-wise soft shrinkage operation as

$$\hat{\mathbf{Z}} = \text{Shrink}_{(N,GM)}(\mathbf{W} \circ \mathbf{X} + \frac{1}{\beta} \lambda_1, \frac{\alpha_1}{\beta}) \quad (22)$$

where (N, GM) is the dimension of shrinkage operation in the objective matrix $\mathbf{W} \circ \mathbf{X} + \frac{1}{\beta} \lambda_1$.

Similarly to \mathbf{Z} -subproblem, the \mathbf{Q} -subproblem also has a closed-form solution:

$$\hat{\mathbf{Q}} = \text{Shrink}_{(N,M)}(\mathbf{W} \circ \mathbf{X} + \frac{1}{\beta} \lambda_2, \frac{\alpha_2}{\beta}) \quad (23)$$

At last, the multipliers matrix λ_1 and λ_2 are updated in standard way as following,

$$\begin{cases} \lambda_1 \leftarrow \lambda_1 - \beta(\mathbf{Z} - \mathbf{W} \circ \mathbf{X}) \\ \lambda_2 \leftarrow \lambda_2 - \beta(\mathbf{Q} - \mathbf{W} \circ \mathbf{X}) \end{cases} \quad (24)$$

Note that there are three variables \mathbf{X} , \mathbf{Z} and \mathbf{Q} updated in our method, which is different from standard ADMM theory in which only two blocks of variables are updated alternatively. However, the functional (16) is separable in \mathbf{Z} and \mathbf{Q} , i.e., $\min_{\mathbf{Z}, \mathbf{Q}} \mathcal{L}(\cdot) = \min_{\mathbf{Z}} \{ \min_{\mathbf{Q}} \mathcal{L}(\cdot) \} = \min_{\mathbf{Q}} \{ \min_{\mathbf{Z}} \mathcal{L}(\cdot) \}$, then the \mathbf{Z} -subproblem and \mathbf{Q} -subproblem in our method can be merged into one subproblem. Hence our method owns the same convergency as the standard ADMM [21]. In short, we have the following result.

Theorem 2. For any $\beta, \alpha_1, \alpha_2 > 0$, and fixed \mathbf{W} , the iteration of \mathbf{X} , \mathbf{Z} and \mathbf{Q} in the ADMM from any initial point is guaranteed to converge to a global minimizer of (15). Specially, \mathbf{X} converges to a solution of (10).

We iterate the ADMM for several times, in each of which we update \mathbf{W} by (14) using the previous estimation of \mathbf{X} as the same as in SOCP. For convenience, we summarize our ADMM in Table I.

TABLE I: ADMM

Algorithm Alternating Direction Method of Multipliers

- 1: Initialize variables \mathbf{Z} , \mathbf{Q} , weight matrix \mathbf{W} , and multipliers λ_1, λ_2 .
 - 2: Initialize parameters $\alpha_1, \alpha_2, \beta, \epsilon$ and the iteration number $MaxCount$. Set the iteration counter $count \leftarrow 1$.
 - 3: **while** $count \leq MaxCount$ **do**
 - 4: **while** not convergent and stopping criterion not met **do**
 - 5: Update \mathbf{X} with (18), \mathbf{Z} with (22), \mathbf{Q} with (23), in turn.
 - 6: Update λ_1 and λ_2 with (24).
 - 7: **end while**
 - 8: Update \mathbf{W} with (14).
 - 9: $count \leftarrow count + 1$.
 - 10: **end while**
-

V. NUMERICAL RESULTS

In this section, we carry out experiments to illustrate the performances of our proposed functional and efficient algorithm. Our SOCP and ADMM for solving functional (10) are addressed as ‘‘SolveSOCP’’ and ‘‘SolveADMM’’ respectively in this simulation. As only the sparsity of user activities is exploited if we eliminate the second term in (10) by setting $\alpha_2 = 0$, we address it ‘‘SolveRowLasso’’ and add it into comparison. Correspondingly, we eliminate the first term in (10) by setting $\alpha_1 = 0$, and address it ‘‘SolveElementLasso’’ which only exploits the sparse relationship between active users and effective RRHs surrounding it. SolveRowLasso and SolveElementLasso can be considered just as special cases of our proposed functional, and both of them are solved by SOCP in this simulation.

The method ‘‘modified Bayesian compressive sensing’’ (‘‘BCS’’ for short) in [9] is also included for comparison. BCS is based on the exact knowledge of large-scale path loss between each RRH and each user, which, however, is costly in real C-RAN and may not be predicted accurately due to mobility and environment changes. Here we include

another case named “BCS2” in which BCS is provided with inaccurate large-scale path loss parameters. We address BCS and BCS2 as “SolveBCS” and “SolveBCS2” respectively in this simulation.

We set the number of RRHs $G = 10$ and number of users $K = 100$. The number of antennas in each RRH and user are $M = 3$ and $N = 2$ respectively. Rayleigh fading channel model is assumed between RRHs and users. The size of active user set \mathcal{A} is 10. SNR is set to 10 dB. Each element of pilot matrix \mathbf{P} in (3) has an independent and identical normalized complex gaussian distribution.

We empirically set α_1 and α_2 in functional (10) equal to a small percentage of α_1^* and α_2^* respectively, say 1%-5%. β in the ADMM functional (15) has the same value level as α_1 and α_2 to accelerate convergence. The number of re-weighting times $MaxCount$ in both SOCP and ADMM is set to 2. ϵ in (14) is set to 10^{-8} . Assuming the precise large-scale path loss between k th user and g th RRH is $\tau_{g,k}$, we provide BCS2 with $\tau_{g,k} \cdot (1 + 0.5n)$ where n obeys standard normal distribution. All the results are averaged over 50 runs.

At first, we demonstrate the channel estimation results of respective algorithms, as shown in Fig. 1 where NMSE denotes the normalized mean squared error. It can be seen that SOCP and BCS provide the best performances. Because of the tradeoff between preciseness and efficiency in the implementation, our ADMM works slightly worse than BCS and SOCP, but still lies in the same level. The performances of “ElementLasso” and “RowLasso” are worse than SOCP which is in accordance with our expectation. BCS2 has a poor performance for reliance on the exact information of large-scale path loss parameters.

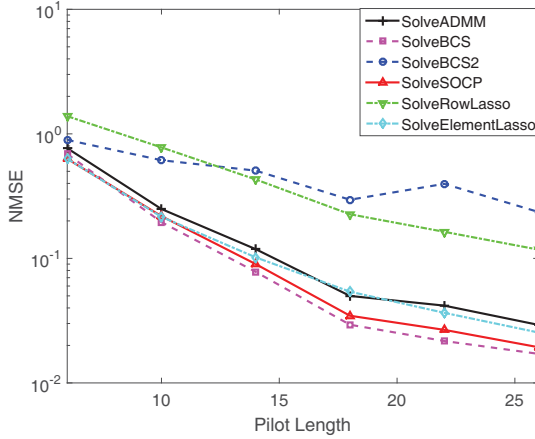


Fig. 1: NMSE of respective algorithms vs. pilot length

Fig. 2 shows the wrongly detected user number of respective algorithms vs. pilot length. We can see that our SOCP and ADMM exhibit the best performances in detecting activities of users. ElementLasso, RowLasso and BCS need longer training pilots to detect users precisely. It can be observed from Fig. 3 that ADMM and BCS have the lowest computational complexity. SOCP has a rather high complexity particularly when pilot is long, which attributes to the increase of problem dimension. The complexity of ADMM and BCS do not increase with pilot length within

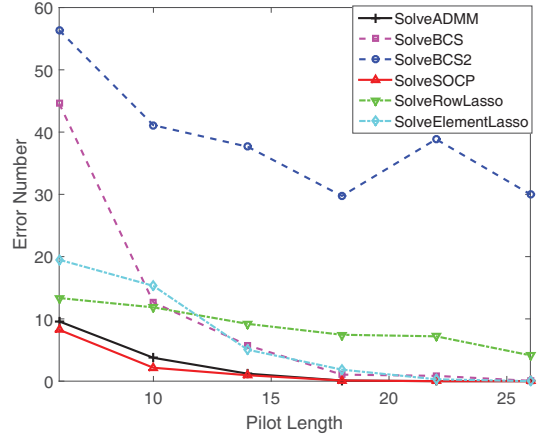


Fig. 2: Multi-user detection results of respective algorithms vs. pilot length

our simulation range because that properly longer pilot helps them to converge faster.

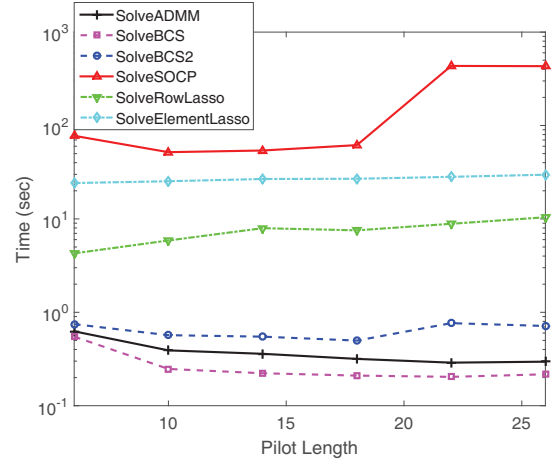


Fig. 3: Computation time of respective algorithms vs. pilot length

VI. CONCLUSION

We proposed a mixed $\ell_{2,1}$ -regularization functional to solve channel estimation and multi-user detection problems in C-RAN. This functional simultaneously exploits the sparsity of user activities and sparsity of RRHs related to each active user. Guidelines are also provided to help choosing tuning parameters. To accelerate the processing procedure, an efficient method based on ADMM is proposed. Numerical results show that our functional solved by our efficient method achieves state-of-the-art performance and also has low computational complexity.

VII. APPENDIX

A. Proof of Theorem 1

At first we prove the solution of (10) $\hat{\mathbf{X}}$ equals to zero if $\alpha_1 \geq \alpha_1^* = \max_i \|(\mathbf{A}_i^H \mathbf{B}) \circ (\mathbf{W}_i)^{\circ(-1)}\|_F$.

To simplify expressions, we denote the three positive terms in (10) as $a_1(\mathbf{X})$, $a_2(\mathbf{X})$ and $c(\mathbf{X})$ in turn. It can be seen that

$$\min_{\mathbf{X}} (a_1(\mathbf{X}) + c(\mathbf{X})) + \min_{\mathbf{X}'} a_2(\mathbf{X}') \leq (10) \quad (25)$$

where equality holds when the minimum solutions $\hat{\mathbf{X}} = \hat{\mathbf{X}}'$.

As $a_1(\mathbf{X}) + c(\mathbf{X})$ is convex, by taking derivative, we get

$$\mathbf{A}^H(\mathbf{A}\mathbf{X} - \mathbf{B}) + \alpha_1 \mathbf{V} = \mathbf{0} \quad (26)$$

where \mathbf{V} is the derivative of $\sum_{i=1}^K \|\mathbf{W}_i \circ \mathbf{X}_i\|_F$ with regard to \mathbf{X} , and each \mathbf{V}_i satisfies

$$\mathbf{V}_i = \begin{cases} \frac{\mathbf{X}_i \circ (\mathbf{W}_i)^{\circ(2)}}{\|\mathbf{X}_i \circ \mathbf{W}_i\|_F} & \text{iff } \mathbf{X}_i \neq \mathbf{0} \\ \in \{\mathbf{V}_i : \|\mathbf{V}_i \circ (\mathbf{W}_i)^{\circ(-1)}\|_F \leq 1\} & \text{iff } \mathbf{X}_i = \mathbf{0} \end{cases} \quad (27)$$

Hence the solution of $\min_{\mathbf{X}} (a_1(\mathbf{X}) + c(\mathbf{X}))$ is zero iff $\|(\mathbf{A}_i^H \mathbf{B}) \circ (\mathbf{W}_i)^{\circ(-1)}\|_F \leq \alpha_1, \forall k = 1, \dots, K$. The solution of $\min_{\mathbf{X}'} a_2(\mathbf{X}')$ is zero. Therefore zero is also the solution of (10) when $\alpha_1 \geq \max_i \|(\mathbf{A}_i^H \mathbf{B}) \circ (\mathbf{W}_i)^{\circ(-1)}\|_F$.

The proof that solution of (10) equals to zero if $\alpha_2 \geq \alpha_2^* = \max_{i,j} \|(\mathbf{A}_i^H \mathbf{B}_j) \circ (\mathbf{W}_{i,j})^{\circ(-1)}\|_F$ can be derived similarly.

B. Complex Matrix Shrinkage

The shrinkage operator for complex matrix is a direct extension of the basic one-dimensional soft thresholding method or shrinkage in, e.g., [22]. To find the solution of an optimization problem as blow

$$\min_{\mathbf{X}} \alpha \|\mathbf{X}\|_F + \frac{\beta}{2} \|\mathbf{X} - \mathbf{B}\|_F^2 \quad (28)$$

where $\mathbf{X}, \mathbf{B} \in \mathbb{C}^{N \times M}$. As (28) is convex but non-differentiable, by taking the derivative of it, we get

$$\alpha \mathbf{V} + \beta(\mathbf{X} - \mathbf{B}) = \mathbf{0} \quad (29)$$

where $\mathbf{V} \in \mathbb{C}^{N \times M}$ is the subgradient of $\|\mathbf{X}\|_F$, and

$$\mathbf{V} = \begin{cases} \frac{\mathbf{X}}{\|\mathbf{X}\|_F} & \text{iff } \mathbf{X} \neq \mathbf{0} \\ \in \{\mathbf{V} : \|\mathbf{V}\|_F \leq 1\} & \text{iff } \mathbf{X} = \mathbf{0} \end{cases} \quad (30)$$

By plugging (30) into (29), we get the optimal solution $\hat{\mathbf{X}} = (\|\mathbf{B}\|_F - \alpha/\beta)\mathbf{B}/\|\mathbf{B}\|_F$ iff $\hat{\mathbf{X}} \neq \mathbf{0}$, and $\|\mathbf{B}\|_F \leq \alpha/\beta$ iff $\hat{\mathbf{X}} = \mathbf{0}$. At last, we arrive at

$$\hat{\mathbf{X}} = \max\{\|\mathbf{B}\|_F - \frac{\alpha}{\beta}, 0\} \frac{\mathbf{B}}{\|\mathbf{B}\|_F} \quad (31)$$

REFERENCES

- [1] China Mobile, "C-RAN: the road towards green RAN," *White Paper*, vol. 2, 2011.
- [2] "ZTE green technology innovations white paper," in *Tech. Report*, 2011.
- [3] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks – a technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, 2015.
- [4] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, 2006.
- [5] H. Zhu and G. B. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Trans. on Commun.*, vol. 59, no. 2, pp. 454–465, 2011.
- [6] B. Shim and B. Song, "Multiuser detection via compressive sensing," *IEEE Commun. Letters*, vol. 16, no. 7, pp. 972–974, 2012.
- [7] H. F. Schepker and A. Dekorsy, "Compressive sensing multi-user detection with block-wise orthogonal least squares," in *Proc. IEEE 75th Veh. Technol. Conf. (VTC Spring)*, Yokohama, Japan, May 2012, pp. 1–5.
- [8] —, "Sparse multi-user detection for cdma transmission using greedy algorithms," in *Proc. 8th Wireless Commun. Syst. (ISWCS)*, Aachen, Germany, Nov. 2011, pp. 291–295.
- [9] X. Xu, X. Rao, and V. K. Lau, "Active user detection and channel estimation in uplink CRAN systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, UK, June 2015, pp. 2727–2732.
- [10] Z. Utkovski, O. Simeone, T. Dimitrova, and P. Popovski, "Random access in C-RAN for user activity detection with limited-capacity fronthaul," *arXiv:1606.02337*, 2016.
- [11] F. Mériiaux, S. Lasaulce, and H. Tembine, "Stochastic differential games and energy-efficient power control," *Dynamic Games and Appl.*, vol. 3, no. 1, pp. 3–23, 2013.
- [12] M. De Mari, R. Couillet, E. C. Strinati, and M. Debbah, "Concurrent data transmissions in green wireless networks: When best send one's packets?" in *Proc. IEEE Int. Symposium on Wireless Commun. Syst. (ISWCS)*, 2012, pp. 596–600.
- [13] D. Har, H. H. Xia, and H. L. Bertoni, "Path-loss prediction model for microcells," *IEEE Trans. on Veh. Tech.*, vol. 48, no. 5, pp. 1453–1462, 1999.
- [14] M. Jeong and B. Lee, "Comparison between path-loss prediction models for wireless telecommunication system design," in *Proc. IEEE Antennas Propag. Society Int. Symposium*, vol. 2, Boston, USA, 2001, pp. 186–189.
- [15] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier anal. Appl.*, vol. 14, no. 5-6, pp. 877–905, 2008.
- [16] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group lasso," *J. Comput. and Graphical Statistics*, vol. 22, no. 2, pp. 231–245, 2013.
- [17] R. Foygel and M. Drton, "Exact block-wise optimization in group lasso and sparse group lasso for linear regression," *arXiv:1010.3320*, 2010.
- [18] L. Meier, S. Van De Geer, and P. Bühlmann, "The group lasso for logistic regression," *J. Royal Statistical Society. Series B*, vol. 70, no. 1, pp. 53–71, 2008.
- [19] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [21] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation*. Upper Saddle River, NJ: Prentice-Hall, 1989.
- [22] N. Parikh, S. P. Boyd et al., "Proximal algorithms," *Found. Trends Optimization*, vol. 1, no. 3, pp. 127–239, 2014.